
Computer Simulation of Membrane Bound Molecules

Abstract of a thesis submitted for the degree of Doctor of Philosophy
in the University of Oxford

Stewart Alan Adcock, Brasenose College, Trinity Term 2001

Transmembrane (TM) proteins form a large and therapeutically important subset of proteins. The physiological roles of TM proteins are varied and they include ion channels, molecular pumps, transporter proteins, receptors, and many others. Around 80 of the top 100 pharmaceuticals act on these TM proteins, illustrating their importance as therapeutic targets. The majority of TM proteins are believed to be members of a structural class known as the helix-bundle proteins. Unfortunately there are significant technical issues which seriously hinder the determination of TM protein structures. Of the 16 000 known protein structures that have been made publicly available only around 160 are TM helix-bundle proteins, and of those only about sixteen are both non-redundant and good quality, which contrasts with estimates that a third of a typical genome encodes TM helix-bundle proteins. Consequently for most cases, any investigation into their molecular mechanisms of activity, or attempts at structure-based drug design, are rendered impossible. Currently, computer generated models are the only way to address this.

The particular properties of the biomembrane environment are important factors for the structure of TM proteins. This prevents most standard protein modelling methods, which were designed for and parameterized using soluble proteins, from being applied in a meaningful way.

Research into the structural analysis, prediction and simulation of TM helix-bundle proteins has culminated in protocols and software to facilitate modelling of the TM domains of such proteins.

As published in the scientific literature, topology prediction, the process by which the TM domains of a given protein are determined from its sequence, may already be solved computationally to a reasonable, but not perfect, level of accuracy. Prediction of the two-dimensional packing arrangement of TM helix-bundles is also reasonably well understood, although this may be open to improvement. Full three-dimensional structure determination is generally intractable except in very limited cases where additional information about the structure is already known.

Novel methods have been successfully applied to build a variety of TM protein models using

software developed in this thesis. Standard comparative protein modelling techniques, homology modelling and threading, may now be performed in a highly automated fashion using scoring functions adapted specifically for the membrane environment. Particularly in the case of the threading technique, previously published scoring functions are not well suited to the unique physical and chemical properties of the membrane environment. A novel hybrid Lamarckian Genetic Algorithm / Monte Carlo procedure was found to effectively pack rigid and flexible helices using a selection of statistically determined empirical scoring functions and atomic level molecular mechanics (MM) energy evaluations along with the application of experimentally determined constraints where appropriate.

The test cases to date include a wide range of structures. In each case, different types of structural information were available. Polyalanine helix-bundles were constructed using simple empirical scoring functions yielding a set of reasonable bundle configurations. A model of the human D₂ dopamine receptor, implicated in several neurological diseases, was built by homology to the crystal structure of bovine rhodopsin generating a structure with the binding site residues correctly placed according to our understanding of the active site, despite not using that information in the construction process. A model of Lactose Permease, a transporter protein from *Escherichia Coli*, was built using a combination of empirical and MM functions with constraints derived from cross-linking experiments.

A range of novel and modified protocols and utilities to aid and automate aspects of TM helix-bundle protein structure modelling and prediction have been designed, implemented and tested. As a set of procedures which are complementary to experimental methodology, they will help to redress the growing fissure between genomic and proteomic research efforts. Using these new tools, a varied range of structural models for a representative set of TM helix-bundle proteins have been successfully constructed, thus proving that modelling of TM proteins is now a feasible task for many realistic problems.